ARMY RESEARCH LABORATORY

Classification of Polar Stratospheric Clouds Using LIDAR Measurements From the SAGE III Ozone Loss and Validation Experiment

by Melvin A. Felton, Jr., Thomas A. Kovacs,
Ali H. Omar, and Chris A. Hostetler

ARL-TR-4154   **June 2007**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Adelphi, MD 20783-1197

ARL-TR-4154   **June 2007**

Classification of Polar Stratospheric Cloudss Using LIDAR Measurements From the SAGE III Ozone Loss and Validations Experiment

Melvin A. Felton, Jr.
Computational and Information Sciences Directorate, ARL

Thomas A. Kovacs
Eastern Michigan University, Ypsilanti, MI

Ali H. Omar and Chris A. Hostetler
NASA LaRC, Hampton, VA

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
June 2007	Final	August 2001 to February 2005

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Classification of Polar Stratospheric Clouds Using LIDAR Measurements From the SAGE III Ozone Loss and Validation Experiment	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Melvin A. Felton, Jr., Thomas A. Kovacs, Ali H. Omar, and Chris A. Hostetler	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Research Laboratory ATTN: AMSRD-ARL-CI-ES 2800 Powder Mill Road Adelphi, MD 20783-1197	ARL-TR-4154

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1197	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Backscatter LIDAR measurements from the stratospheric aerosol and gas experiment (SAGE) III ozone loss and validation experiment (SOLVE) have been used to identify classes of polar stratospheric clouds (PSCs) and their corresponding characteristics. Volume backscatter at 532 nm (β_{532}) and 1064 nm (β_{1064}), scattering ratio at 532 nm (R_{532}) and 1064 nm (R_{1064}), aerosol depolarization (δ_a), and color ratio (β_{532}/β_{1064}), and temperature (T) were used to categorize PSCs via principal component analysis (PCA) and Cluster Analysis (CA). The clusters found in this study are consistent with previous studies of PSCs of types Ia ($R_{532} < 1.5$ and $\delta_a > 0.1$), Ia-enhanced ($R_{532} > 1.4$ and $\delta_a > 0.1$), Ib ($R_{532} > 2$ and $\delta_a < 0.025$), and II ($R_{532} > 7$ and $\delta_a > 0.1$). In addition, a cluster predominantly found on the outside edges of clouds, which does not fit the definition of these PSCs, which has low R_{532} and δ_a values, was found. The clustering analysis of PSC data provides an objective and automated method for the production of image plots that show a two-dimensional distribution of the PSC particle types. Such capabilities will be useful for analyzing large data sets such as will be produced by space-borne LIDARS (laser identification and ranging system).

15. SUBJECT TERMS

Polar stratospheric clouds, cluster analysis

16. Security Classification of:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Melvin A. Felton, Jr.
a. REPORT	b. ABSTRACT	c. THIS PAGE	U	36	19b. TELEPHONE NUMBER *(Include area code)* (301) 394-2618
Unclassified	Unclassified	Unclassified			

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

1. Introduction

The National Research Council (NRC) lists stratospheric ozone and the species that control its catalytic destruction as a key research challenge facing the atmospheric chemistry community in the 21st century (NRC, 1998). The physical properties of aerosols and clouds affect stratospheric chemistry in general and stratospheric ozone in particular. Because chemical reactions that take place on the surfaces of aerosols and polar stratospheric clouds (PSCs) have been connected to the observed springtime ozone depletion in both the Arctic and Antarctic, it is important to know PSC phase, size, and chemical composition. Laboratory studies have shown that the heterogeneous reactions that occur on the different types of PSCs are all effective for activating chlorine but with different efficiencies (Carslaw et al., 1997). The ability to accurately classify PSCs according to their physical characteristics will assist in our studying their reactivity. In addition, using an automated method to classify PSCs as opposed to the more subjective approaches used in the past can result in a more efficient method for analyzing large data sets, particularly those from space-borne LIDARS (laser identification and ranging system).

LIDARS have been used to study PSCs in the Arctic and Antarctic (McCormick et al., 1981, Poole & McCormick, 1988; Kent et al., 1990; Browell et al., 1990; Toon et al., 1990; Tabazadeh & Toon, 1996; Tsias et al., 1999; Toon et al., 2000).

These studies have yielded many PSC particle classifications that were (at least initially) based on apparent clustering of LIDAR observables: scattering ratio, aerosol depolarization, and color ratio. The boundaries between the clusters were defined somewhat subjectively by scientists analyzing the data and were subsequently corroborated by deductions of microphysical distinctions (particle size, shape, and composition) suggested by the LIDAR observables and the temperatures at which the PSCs were observed. The most common of these classifications is presented in table 1.

Table 1. Summary of PSC particle classifications.

PSC Classification	Description
Ia	Large nitric acid trihydrate crystals
Ia-enhanced	NAT crystals down wind from mountain wave-induced ice clouds
Ib	Liquid ternary solution particles
Ic	Ia and Ib mix
II	Ice crystals

In this study, the extensive set of PSC measurements was acquired during the stratospheric aerosol and gas experiment (SAGE) III ozone loss and validation experiment (SOLVE) and was used as the basis to classify PSCs. Well-known statistical techniques were employed in an

attempt to more objectively identify the number of PSC types and their corresponding characteristics.

This technique is in contrast to many of the methods used in the past which used a more subjective approach that relied on the researchers' ability to discriminate between particle types, based upon their knowledge of the parameters used in their analysis. The method presented here is potentially a more efficient method for classifying and analyzing the PSCs within the extensive data sets that are produced by space-based LIDAR missions such as the cloud-aerosol LIDAR and infrared pathfinder satellite observations (CALIPSO).

2. SOLVE Aerosol LIDAR Data

SOLVE was a measurement campaign focused on the processes that control ozone concentrations at middle and high latitudes. The mission included the deployment of several aircraft-based and balloon-borne instruments and was based in Kiruna, Sweden, during the winter of 1999-2000. This study employed data acquired by the LaRC aerosol LIDAR, which was deployed on the National Aeronautics and Space Administration (NASA) DC-8 aircraft on the SOLVE mission. The LaRC aerosol LIDAR is a piggy-back instrument on the NASA Goddard Space Flight Center (GSFC) airborner Raman ozone, temperature, and aerosol LIDAR (AROTAL). The LIDAR data used in this study were limited to the following: total scattering ratio at 532 nm (R_{532}), total scattering ratio at 1064 nm (R_{1064}), aerosol backscatter coefficient at 532 ($\beta532$), aerosol backscatter coefficient at 1064 nm (β_{1064}), and aerosol depolarization ratio at 532 nm (δ_{532}). Each measurement in the data set pertains to an observed ensemble of particles.

The backscatter ratio provides information about the ratio of the backscatter from aerosols to the backscatter from molecules, and the aerosol backscatter coefficient is a function of the concentration and size of the aerosols. Therefore, for an ensemble, a high scattering ratio implies that there is more backscatter attributable to aerosols than to molecules, and a high aerosol backscatter coefficient implies a high concentration of aerosols or the presence of large aerosols. Also, information about the shapes of the particles in the ensemble can be obtained from the aerosol depolarization ratio. A high aerosol depolarization ratio indicates that non-spherical particles are a part of the ensemble being measured which is associated with solid particles as opposed to ensembles of liquid particles that produce low aerosol depolarization ratios. The color ratio used in this analysis is defined as β_{532}/β_{1064} and provides information about the sizes of the particles in the ensemble attributable to the wavelength dependence for particle scattering. Therefore, a high color ratio indicates the presence of small particles within the ensemble and a small color ratio indicates the presence of large particles within the ensemble. The vertical and horizontal resolutions of the data products used in this study were 75 m and 2.5 km, respectively.

Filters were applied to the SOLVE LIDAR data to screen the PSC data for use in the cluster analysis. Only data acquired during nighttime lighting conditions were used to ensure the greatest signal-to-noise ratio and thus the lowest chance of misidentifying a noise excursion as a PSC. The altitude range was restricted to 14 to 26 km, the lower limit being the lowest range of the LIDAR data and the upper limit set at the maximum altitude at which PSCs were observed during the SOLVE mission. The scattering ratios were restricted to $R_{532} > 1.12$ and $R_{1064} > 1.6$ to isolate PSC data for use in the analysis, and only temperatures below 198 K were included. The final number of data points satisfying these filters and used as objects in the cluster analysis is 18,275.

During the SOLVE mission, there were occasions when strong signals from PSCs saturated the detection system (i.e., the signal level exceeded the range of an amplifier or analog-to-digital converter), making the measurements inaccurate. Unfortunately, it is not possible in post-flight analysis to determine which data were corrupted by saturation, and this may have affected some of the results presented in this report.

3. Method

3.1 Principal Component Analysis

Principal component analysis (PCA) is a statistical method used to resolve the complicated variance of the multivariate set of data consisting of β_{532}, β_{1064}, R_{532}, R_{1064}, δ_a, β_{532}/β_{1064}, and temperature (T). Temperature profiles were derived from the global modeling and assimilation office (GMAO) meteorological model. The analysis was done with the standardized values of the variables, which makes them equally important by creating new variables that each have a mean of zero and a variance of one. The PCA essentially redefines the data set in terms of derived variables that are based on linear combinations of the original variables (Preisendorfer, 1988). Each data point in the derived variable data set corresponds to exactly one data point in the original data set.

The weight assigned to each derived variable is also a part of the output of PCA. Thus, the derived variables that account for large portions of the variance of the data can be identified, thereby reducing the dimensions of the analysis by disregarding the derived variables that account for a minimal variance. Consequently, information concerning the variance of each original variable may also be obtained. A more detailed description of the method is provided by Felton (2003).

3.2 Cluster Analysis

A clustering algorithm was used to classify the clouds into groups, based on combinations of the derived variables from the PCA. A particular partitioning method has been chosen to perform

3

the cluster analysis. The technique is based on the search for representative objects among the many objects of the data set called "medoids" (Struyf et al., 1996). The medoids are calculated so that the total dissimilarity of all objects to their nearest medoid is minimal. A range of values for the number of clusters (k) desired is required as input for the clustering algorithm. The natural number of clusters can be obtained from analysis of a quality index calculated for each cluster as well as the corresponding graphical output (Kaufman & Rousseeuw, 1990). This index, named the "silhouette coefficient" (SC) by the authors, provides an indication of the relationship between the objects of a cluster. The authors' suggested interpretation of the values of the silhouette coefficient is that 0.71 to 1.00 imply that the clusters are well defined; 0.51 to 0.71 imply that the clusters are reasonably defined; 0.26 to 0.50 imply that the clusters are poorly defined; and ≤ 0.25 imply that no substantial groupings have been found.

Note that it is possible to obtain a relatively high silhouette coefficient for a value of k that is not the most natural number of clusters for the data set.

For this reason, Kaufman and Rousseeuw (1990) suggested that the value of k that yields the highest silhouette coefficient should not be selected as the natural number of clusters for the data set unless the graphical output is also examined. The groups should be such that the degree of association is strong between members of the same cluster and weak between members of different clusters. An example of this partitioning method is shown in figure 1.
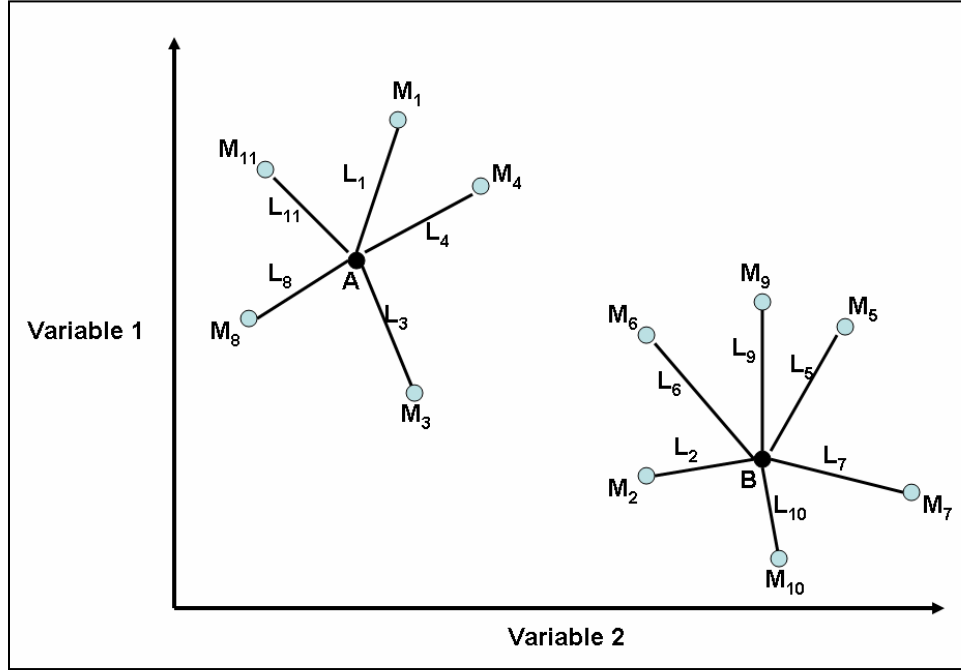
Figure 1. An example of the partitioning method where the M_i are data points used in the analysis. (A and B denote the medoids of the clusters. L_i are distances from medoids to members of the cluster. The algorithm minimizes this distance. This example is a 2-D representation of the clustering space.)

3.3 Simulation

To test the clustering algorithm, random numbers were generated that are normally distributed about the means for the R_{532} and aerosol depolarization ratios of the PSC types defined by Browell et al. (1990). Thus, the test data were modeled after types Ia, Ib, and II PSCs. There were an equal number of points for each PSC type, and nonphysical points were not included in the simulation because they were excluded from the data set altogether. The algorithm was asked to partition the data into two to five clusters (k = 2…5). The means and standard deviations used to generate the data are given in table 2. The resulting silhouette coefficients for the partitioned test data are reported in table 3. As suggested by Kaufman and Rousseeuw (1990), the silhouette coefficients and the actual graphical results of the clustered data should be used to assess the validity of the classification.

5

Table 2. Mean (μ) and standard deviations (σ) for R_{532} and δ_a used to generate the test data.

Type	$\mu_{R_{532}}$	$\sigma_{R_{532}}$	μ_{δ_a}	σ_{δ_a}
Ia	1.32	0.20	0.400	0.100
Ib	5.00	1.00	0.015	0.010
II	15.00	3.50	0.200	0.035

Table 3. Mean values, μ, and standard deviations, σ, for R_{532} and δ_a for the three clusters resulting from the k = 3 partition of the test data.

Cluster	$\mu_{R_{532}}$	$\sigma_{R_{532}}$	μ_{δ_a}	σ_{δ_a}
black	1.414	0.426	0.403	0.097
red	4.975	1.177	0.022	0.028
yellow	15.035	3.257	0.199	0.034

The cluster plots for k = 2…5 are shown in figures 2 a, b, c, and d. In the case of k = 2, one cluster is formed as the union of the type Ib and type II PSCs. Although the silhouette coefficient of 0.57 suggests that these may be reasonably defined clusters, we know that type Ib and type II PSCs are distinct from each other because of their depolarization values. When k = 3, all three PSC types are accurately identified. Despite outlying points having a negative effect on the corresponding silhouette coefficient, a value of 0.73 suggests that strong clusters have been found. For k = 4, the type Ia and Ib PSCs are accurately identified but the type II PSCs are partitioned into two clusters. The lack of dissimilarity between the two clusters resulting from the partitioning of the type II PSCs results in a lower silhouette coefficient than the 0.73 value for k = 3. When k = 5, only the type Ib PSCs are accurately identified as a cluster. The type II and type Ia PSCs are partitioned into two clusters each. The corresponding silhouette coefficient decreases to 0.62 because of the lack of separation between the clusters resulting from the partitioning of the type Ia and type II PSCs.
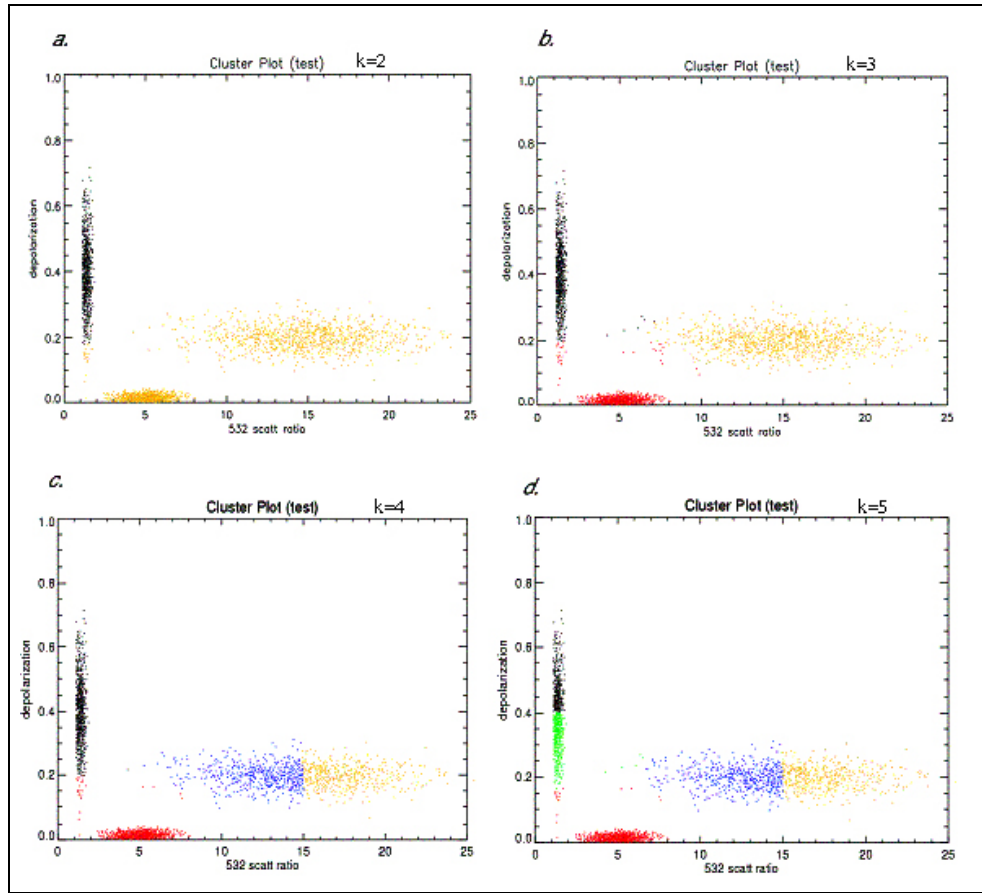
Figure 2. Test data partitioned into k = 2…5 clusters (a. k = 2 where the yellow cluster is the union of type Ib and type II and the black cluster is type Ia. b. k = 3 where the black cluster is type Ia, the red is type Ib, and the yellow is type II. c. k = 4 where the black cluster is type Ia, the red is type Ib, and the blue and yellow represent the type II split into two separate clusters. d. k = 5 where the red is type Ib, the black and green represent type Ia split into two separate clusters, and the blue and yellow represent type II split into two separate clusters.)

Thus, the silhouette coefficient of 0.73 suggests that three is the most natural number of clusters for the test data set. When we look at the corresponding cluster plot, we see that all three types of PSCs included in the test are accurately identified as separate clusters. Also, from examining the cluster plots of the other values of k, we see that the silhouette coefficient is a function of the homogeneity of the clusters as well as their relative dissimilarities to each other. Table 3 shows the mean values and standard deviations of the clusters resulting from the k = 3 partition. We can also see that the algorithm has successfully identified the type Ia, Ib, and II PSCs (black, red, and yellow clusters, respectively) by comparing the resulting values in table 3 to the values used to generate the test data in table 2. However, although the yellow and black clusters are very homogeneous, the red cluster does contain some data points that do not have Ib characteristics. For instance, type Ia particles with low depolarization are included in the red cluster. In addition, particles with a mixture of Type Ia and Ib properties ($R \approx 6$ and $\delta \approx 0.15$) are included in the red cluster.

7

4. Results

Different combinations of R_{532}, R_{1064}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}, and T were used as input for the analysis to see which combination of variables is best for identifying the different types of PSCs. The variable combinations that included at least three of the four scattering variables (R_{532}, R_{1064}, β_{532}, and β_{1064}) resulted in the highest separation between clusters. These variable combinations are shown in table 4. The results of performing the analysis on variable combinations with just scattering ratio or backscatter coefficient are presented in Felton (2003). These variable combinations resulted in clusters that have poor separation with respect to each other. In all cases, PCA was used to reduce the dimensions of the analysis while still capturing at least 95% of the total variance of the data set. The results are presented in three sections. The first section consists of the results of the analysis performed on data from all 11 SOLVE flights for which acceptable data were available. Analyzing large amounts of data in this manner shows the method's usefulness in the processing of satellite data in real time for a first order classification of PSC types. The second and third sections focus on individual days that have a significant difference in their lowest temperatures.

Table 4.   The different variable combinations that have been used in the analyses.

Combination No.	Variables
A	R_{532}, R_{1064}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}, T
B	R_{532}, R_{1064}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}
C	R_{532}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}, T

4.1   Full Data Set

The percent of the variance attributed to each derived variable used in the analysis for variable combinations A, B, and C is shown in table 5. For each combination, the main contributors to derived variables 1, 2, and 3 are the scattering variables, aerosol depolarization and color ratio, and temperature, respectively. Therefore, for A, the scattering variables account for approximately 56% of the variance while aerosol depolarization and color ratio accounts for approximately 24% of the variance. When the PCA was performed on B, the variance accounted for by the scattering variables increased by roughly 7% and the variance accounted for by aerosol depolarization and color ratio increased by less than 4%. For C, the variance accounted for by the scattering variables decreased to 50% of the total variance and that of aerosol depolarization, color ratio, and temperature increased slightly.

8

Table 5.   The percentage variance of the data set attributed to each derived variable used in the analyses for variable combination A, B, and C.

Derived Variable	Percent variance for A	Percent variance for B	Percent variance for C
1	56	63	50
2	24	28	28
3	12	5.9	14
4	5.0	N/A	5.8

Comparisons of the analysis performed on variable combinations A, B, and C are shown in figures 3 and 4. The presence of liquid or solid particles within the ensembles can be inferred from R_{532} versus δ_a plots of figure 3, and the presence of large or small particles within the ensembles can be inferred from the R_{532} versus β_{532}/β_{1064} plots of figure 4. For each variable combination, the data set was partitioned into k = 2, 3, …6 clusters and the partitions corresponding to the three highest silhouette coefficients are presented.
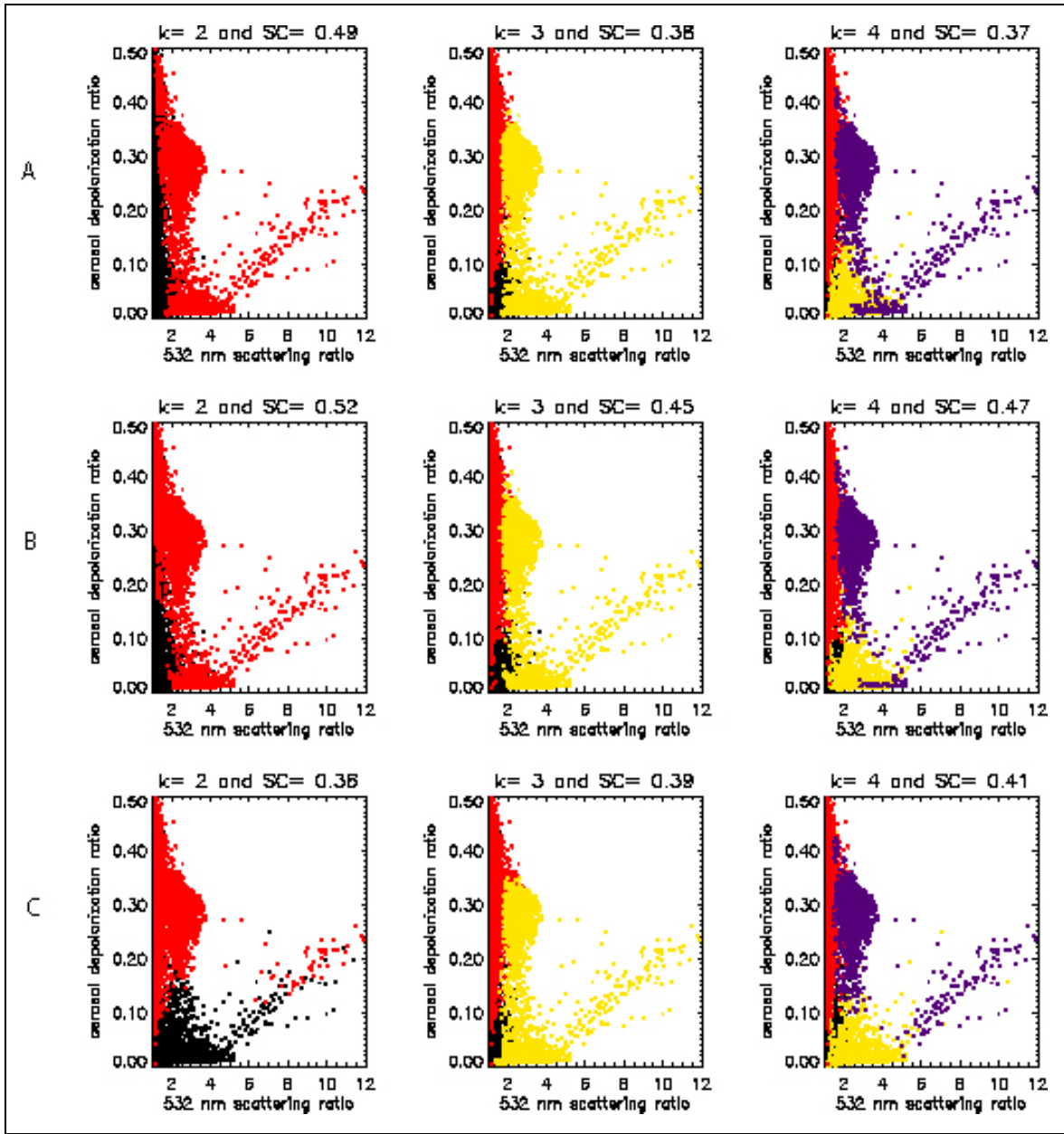
Figure 3. Comparisons of R_{532} versus δ_a for the analysis performed on variable combinations A, B, and C.
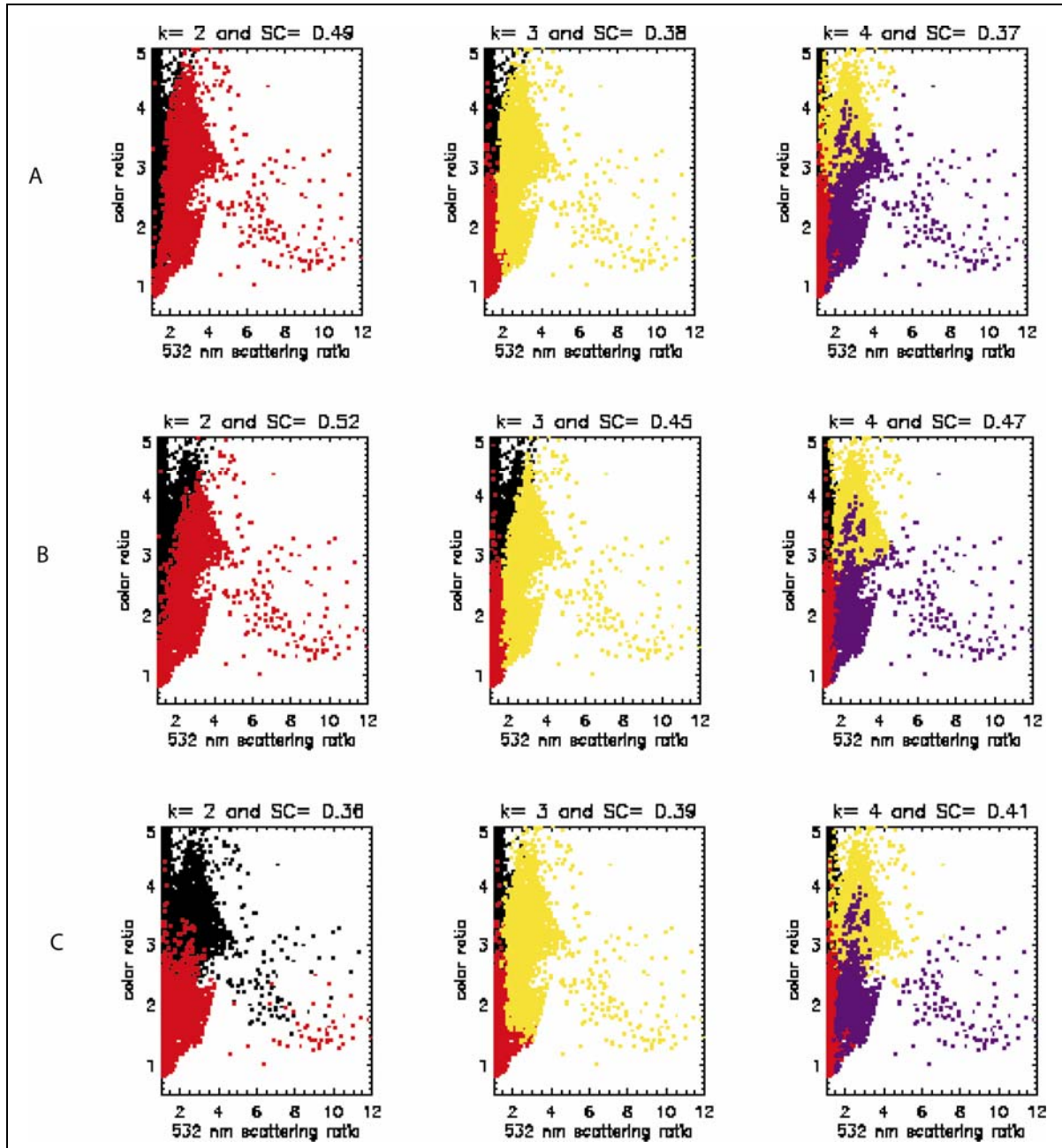
Figure 4. Comparisons of R_{532} versus β_{532}/β_{1064} for the analysis performed on variable combinations A, B, and C.

The clusters are represented by color.

For A, in figures 3 and 4, the SC of 0.49 suggests that k = 2 is the best partition for the data set. Since all of the scattering variables were used, particle scattering dominates the partitioning and separates low and high scattering measurements indicated by the black and red clusters, respectively. The clusters contain measurements with depolarizing and non-depolarizing particles suggesting that they are not entirely homogeneous. When k = 3, the low scattering measurements are separated according to depolarization indicated by red and black clusters and a

yellow cluster continues to consist of non-depolarizing and depolarizing measurements. When k = 4, the yellow cluster becomes entirely non-depolarizing measurements and a purple cluster emerges that is poorly defined, consisting of outlying measurements with very high R_{532} that are non-depolarizing with high color ratios or depolarizing with smaller color ratios.

The results for B in figures 3 and 4 are almost identical to those of A. Eliminating T from the analysis resulted in increased SCs, which suggests that many of the clusters may occur at similar temperatures. This combination of variables also reduces the amount of measurements with small non-depolarizing particles in the purple cluster for k = 4.

For C in figures 3 and 4, eliminating R_{1064} adds more weight to δ_a. When k = 2, the data are now partitioned with respect to δ_a in addition to R_{532}. The clusters that result from the k = 3 partition remain the same as the three clusters in A and B. The highest SC occurs when k = 4 and the clusters are now very homogeneous except for the purple cluster, which still contains the outlying measurements containing large particles with high R_{532}.

Variable combination C results in clusters that best represent the known PSC types. For variable combinations A and B (which include R_{1064} in the analysis) in figures 3 and 4, the k = 4 partition results in a purple cluster that contains measurements of ensembles with large particles centered at $\delta_a = 0.3$ as well as few small non-depolarizing particles. Although many of these depolarizing measurements have lower R_{532} values than the non-depolarizing measurements, both have similar R_{1064} values because the depolarizing measurements contain large particles and are more efficient at scattering at 1064 nm. Therefore, including R_{1064} impedes the accurate classification of the non-depolarizing measurements with high scattering ratio values. Therefore, variable combination C will be used in the analyses of the individual days in the following sections.

The black cluster that results from the use of variable combination C to partition the data set into four clusters (figures 3 and 4) consists of measurements of ensembles with small particles and has mean values of $R_{532} = 1.252$ and $\delta_a = 0.041$. These particles are centered around the origin and are most likely the aerosols that are precursors to PSCs. The red cluster resembles type Ia measurements with mean values of $R_{532} = 1.235$ and $\delta_a = 0.231$ and the yellow cluster resembles type Ib measurements having mean values of $R_{532} = 2.503$ and $\delta_a = 0.024$. Most of the purple cluster, centered at $R_{532} = 2.5$ and $\delta_a = 0.3$, consists of measurements that resemble type Ia-enhanced. Included in this cluster are measurements that have very high R_{532} as do the type II PSCs.

Not shown in figures 3 and 4 are the k > 4 partitions where the algorithm partitions other clusters further before successfully identifying the measurements indicative of the type II particles. Two factors prevent the algorithm from successfully identifying these measurements. First, type II measurements are far less numerous than other types of PSC measurements and tend to look like outliers of the Ia-enhanced cluster. The filtered data set has a total of 18,275 data points while the number of data points with $R_{532} > 5$ is 176. Secondly, the type II PSC measurements share optical characteristics such as high δ_a and low β_{532}/β_{1064} with the type Ia enhanced.

4.2 CA For January 23, 2000

The cluster algorithm partitioned the 6,089 measurements (objects) from January 23, 2000, into k = 2…6 clusters. A flight map is provided in figure 5. Some of the measurements on this day occurred during periods when the temperature was less than 188 K.
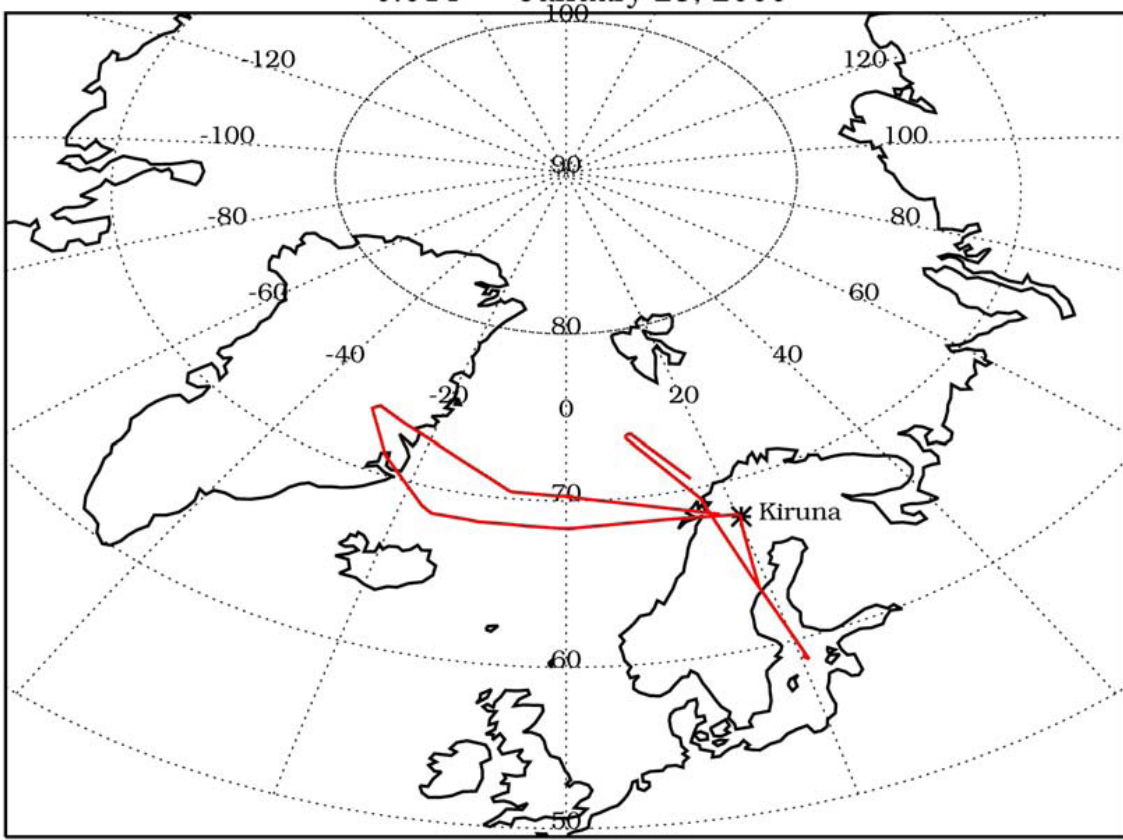
Figure 5. January 2, 2000 flight map.

Variable combination C was used in the analysis, and the resulting SCs are shown in table 6. Figure 6 shows R_{532} versus δ_a and figure 7 shows R_{532} versus β_{532}/β_{1064} for the three highest values of k where the blue stars are the representative objects (medoids) for each cluster. The two clusters resulting from the k = 2 partition are not representative of the number of PSC types present because they are composed of measurements of ensembles with small and large particles as well as liquid and solid particles. When k = 4, the SC reaches its maximum value of 0.40. A black cluster with low R_{532} and low to moderate δ_a and a yellow cluster with low to moderate R_{532} and low δ_a are found. Other resulting clusters include a red cluster with low to moderate R_{532}, moderate to high δ_a, and moderate to high β_{532}/β_{1064} and a purple cluster with high R_{532} and δ_a, and low to moderate β_{532}/β_{1064}. There is good separation between the medoids of these clusters except for the yellow and black clusters. Partitioning the data set five times results in a decline in SC to 0.35 because of the lack of separation between the red and green clusters.

13

Table 6. Silhouette coefficients for 2 to 6 clusters of the January 23 data set.

No. of clusters	SC
2	0.38
3	0.33
4	0.40
5	0.35
6	0.32
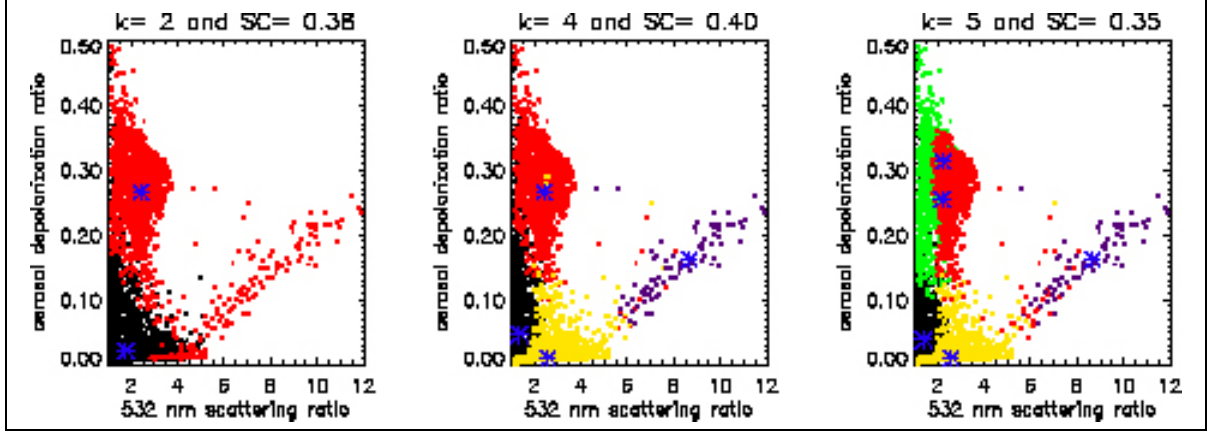
Figure 6. R_{532} versus δ_a for k = 2, 4, and 5 for the January 23 data set. (The blue stars are the representative objects of the clusters.)
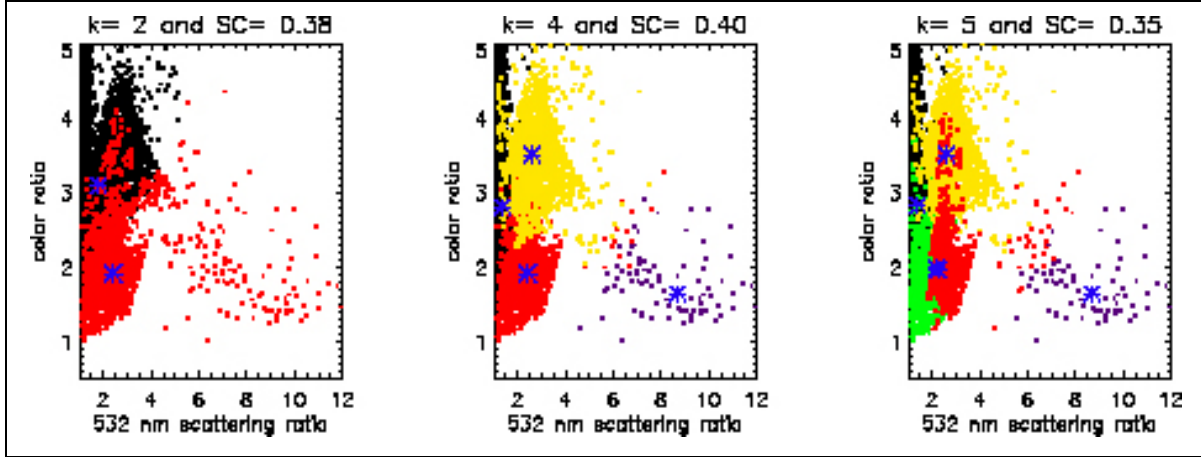
Figure 7. R_{532} versus β_{532}/β_{1064} for k = 2, 4, and 5 for the January 23 data set. (The blue stars are the representative objects of the clusters.)

Therefore, the optimum number of clusters for January 23, 2000, is four. The mean values of these four clusters are shown in table 7. Despite having low R_{532}, the black cluster's low δ_a suggests that not all of the particles in these ensembles are the type Ia described by Browell et al. (1990). From figure 6, it can be seen that there are many measurements in the black cluster with $\delta_a < 5\%$, which suggests the presence of more liquid particles than solid type Ia particles. Thus,

the black cluster seems to be composed of a mixture of type Ia and precursor aerosol particles. The red cluster matches very well with the type Ia-enh described by Tsias et al. (1999) and Reichardt et al. (2000) which has both moderate R_{532} and δ_a. The measurements of the yellow cluster that contain small spherical particles are predominantly the type Ib described by Browell et al. (1990).

Table 7. Mean values of R_{532}, δ_a, and β_{532}/β_{1064} for k = 4 clusters for January 23 data set.

Cluster	Mean R_{532}	Mean δ_a	Mean β_{532}/β_{1064}	Mean T
Black	1.386	0.064	2.962	192
Red	2.365	0.284	1.985	191
Yellow	2.791	0.024	3.492	188
Purple	12.563	0.190	1.844	184

The purple cluster is the type II PSCs described by Browell et al. (1990) with its high R_{532} and moderate δ_a. This PSC type was not identified as a separate cluster when the entire data set was used as input to the analysis. Using the smaller number of data points in the January 23 data set made it possible to cluster this class separately, probably because type II PSCs represent a larger fraction of the observations from the flight of 23 January than they do in the composite of observations from all the flights.

Figure 8 is an image plot for January 23, 2000, which illustrates the spatial proximity of the different types of PSCs in the CA. The first three panels in the figure show R_{532}, δ_a, and β_{532}/β_{1064}, and the last panel shows the classification of the PSC measurements resulting from the CA. The type Ia and precursor mixture, type Ia-enh, type Ib, and type II correspond to black, red, yellow and purple, respectively. On this particular day, the aerosols were found along the outer edges of the type Ia-enhanced, Ib, and II PSCs, and the type Ib and type II PSCs were found adjacent to each other in the coldest regions of the cloud.

Figure 8.   Image plots for January 23, 2000. (Panels a, b, and c show R_{532}, δ_a, and β_{532}/β_{1064}, respectively, and panel d shows the classification of the LIDAR data into clusters. The colors are as follow: black- PSC type Ia and precursor aerosol mixture, red- PSC type Ia-enh yellow- PSC type Ib, and purple- PSC type II. The gap in d represents data filtered out of the analysis because of excessive noise from high solar background light.)

4.3   CA For March 5, 2000

The cluster algorithm was used to partition the 3,844 measurements from March 5, 2000 into k = 2…4 clusters. The flight map for this day is shown in figure 9. The temperatures corresponding to the measurements on this day were above 188 K. The resulting silhouette coefficients and the graphical output are shown in figures 10 and 11.
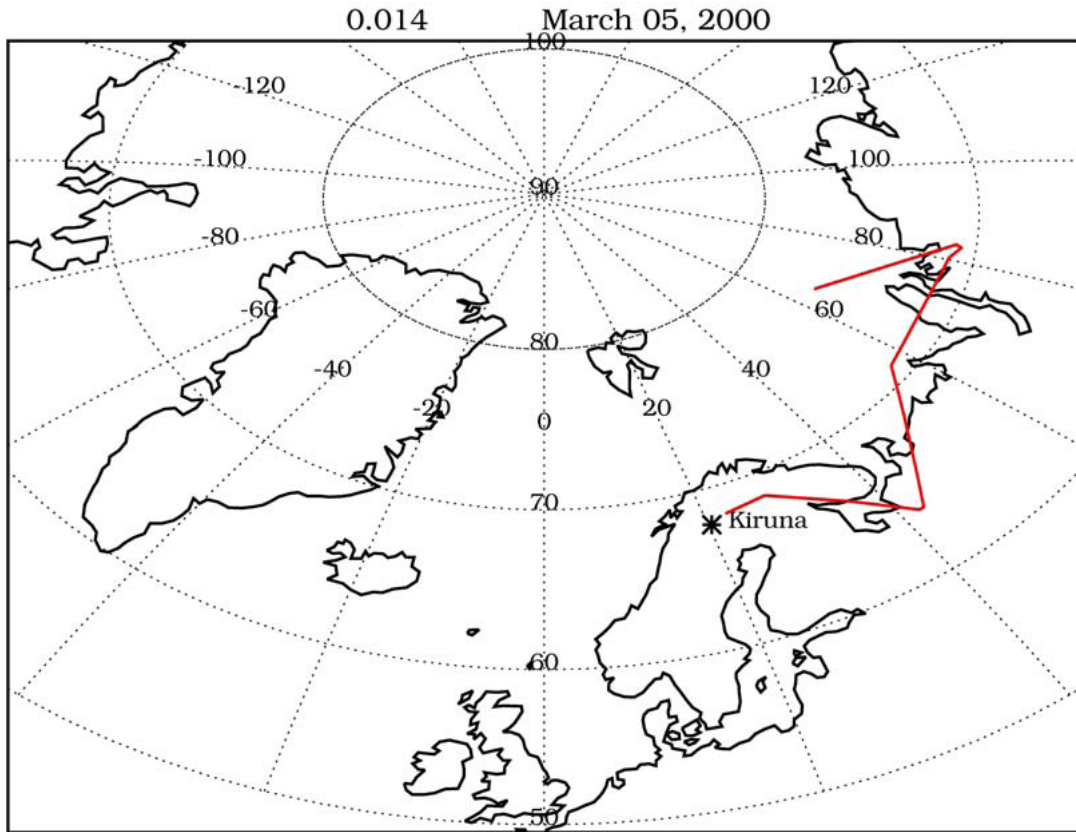
Figure 9. March 5, 2000 flight map.
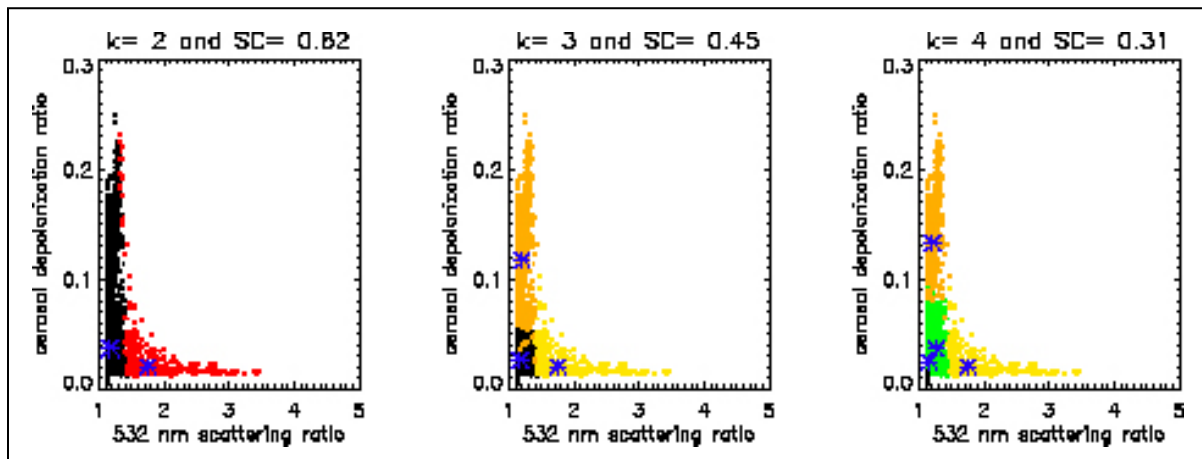
Figure 10. R_{532} versus δ_a for k = 2, 3, and 4 for March 5 data set. (The blue stars are the representative objects of the clusters.)
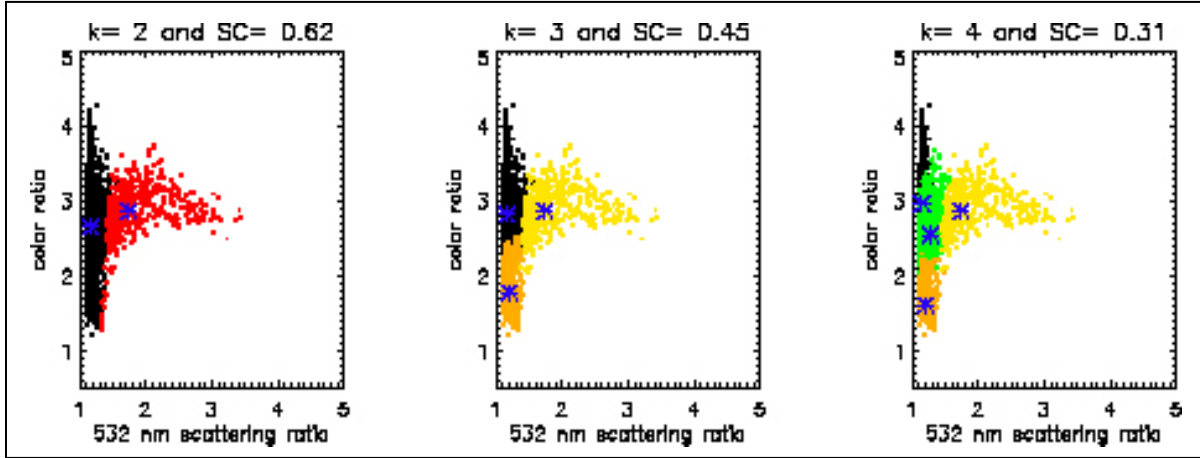
Figure 11. R_{532} versus β_{532}/β_{1064} for k = 2, 3, and 4 for March 5 data set. (The blue stars are the representative objects of the clusters.)

Although the two clusters (k = 2) result in the highest SC, the range of depolarization ratios suggests that the black cluster must consist of measurements of ensembles with solid and liquid particles. When k = 3, the SC drops to 0.45 because while the yellow cluster remains the same, the black cluster from k = 2 has been partitioned into two clusters with similar R_{532} represented by the orange and black clusters. Despite the lower SC, the k = 3 partition more accurately describes the particles present because the measurements with liquid particles and the measurements with solid particles, represented by the black and orange clusters, respectively, are now classified as separate clusters. When k = 4, the SC decreases to 0.31 because of the extraneous partition resulting in the green cluster whose representative object is very similar to that of the black cluster. Three is therefore the most appropriate number of classes for the PSCs observed on March 5, 2000. The mean values of these three clusters are shown in table 8.

Table 8. Mean values of R_{532}, δ_a, and β_{532}/β_{1064} for k = 3 clusters of March 5 data set.

Cluster	Mean R_{532}	Mean δ_a	Mean β_{532}/β_{1064}	Mean T
Black	1.189	0.034	2.841	194
Orange	1.211	0.129	1.847	194
Yellow	1.879	0.023	2.914	193

The black cluster has optical characteristics very similar to the black cluster in the data set for January 23, 2000. This cluster is likely a mixture of solid and liquid aerosol particles. The mean values for the orange cluster are almost identical to the type Ia found by Browell et al. (1990) with $R_{532} < 1.5$ and $\delta_a > 0.1$. The presence of the small spherical particles of the yellow cluster is consistent with type Ib particles found by Browell et al. (1990).

Image plots for March 5, 2000, are shown in figure 12. The first three panels in the figure show R_{532}, δ_a, and β_{532}/β_{1064}, and the last panel shows the classification of the PSC particles resulting

18

from the CA. The type Ia and precursor mixture, type Ia, and type Ib correspond to black, orange, and yellow, respectively. The aerosols are found on the outer edges of the other PSC types. This cloud displays a layer of type Ib PSCs directly above a layer of type Ia PSCs.
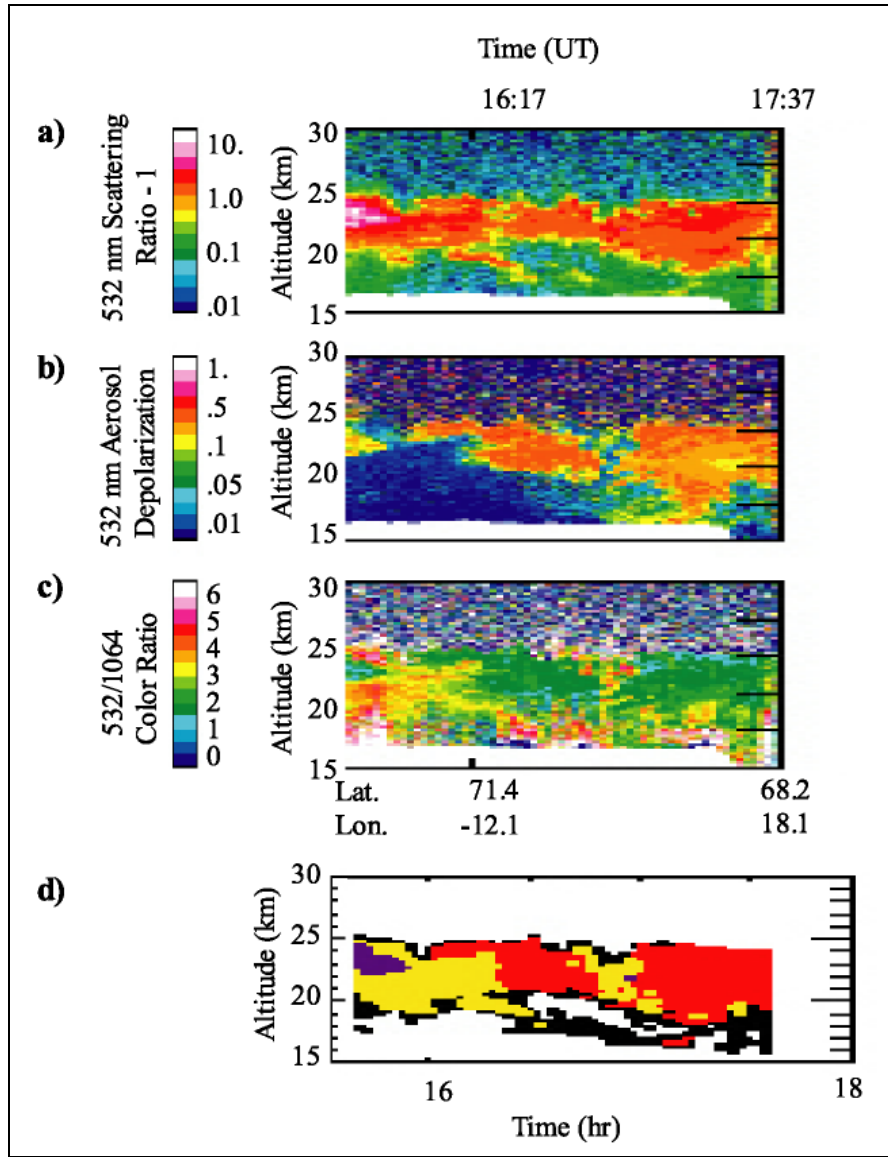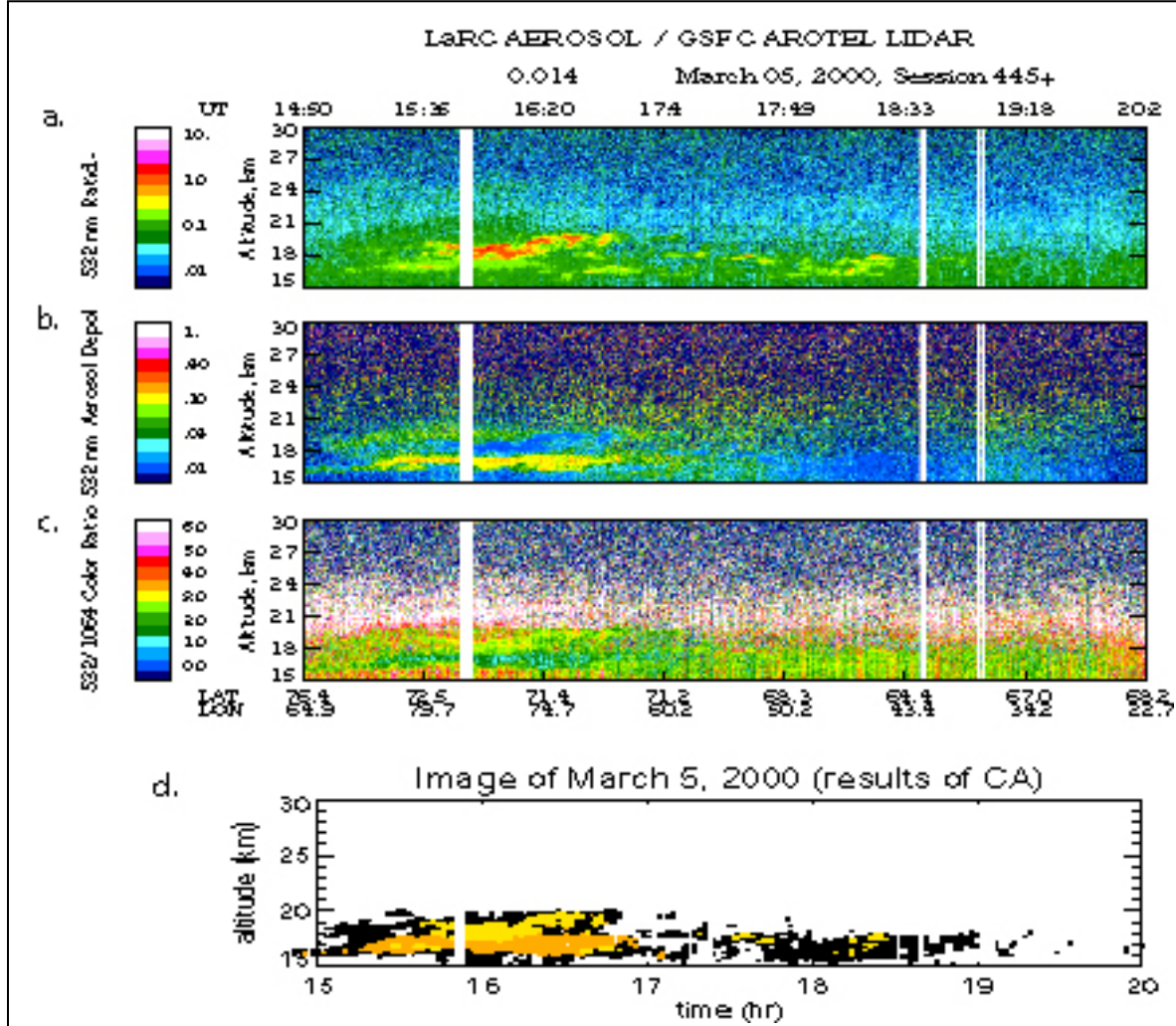

Figure 12. Image plots for March 5, 2000. (Panels a, b, and c show R_{532}, δ_a, and β_{532}/β_{1064}, respectively, and panel d shows the classification of the LIDAR data into clusters. The colors are as follows: black- PSC type Ia and precursor aerosol mixture, yellow- PSC type Ib, and orange- PSC type Ia.)

4.4 Clusters in Analyses

A total of five particle types has been identified in the cluster analyses of the SOLVE LIDAR data. Four of the clusters are known PSC types and the remaining cluster resembles a mixture of liquid and solid background aerosol particles that are precursors to the PSC particles. This precursor cluster, identified as the black cluster in this study, was often found on the outside edges of clouds as in figures 8 and 12 and is similar to the findings of Biele et al. (2001). All five

PSC types were observed on January 23 and March 5 collectively. On the other nine days, different combinations of some of these five clusters, summarized in table 9, were observed.

Table 9. Classification and mean characteristics of the clusters found on January 23 and March 5.

Cluster	Classification	Mean R_{532}	Mean δ_a	Mean β_{532}/β_{1064}	Mean T
Black	Liquid/solid aerosol mixture	1.288	0.049	2.902	193.001
Orange	Ia	1.211	0.129	1.847	194.097
Red	Ia-enhanced	2.365	0.284	1.985	190.929
Yellow	Ib	2.335	0.024	3.203	190.795
Purple	II	12.563	0.190	1.844	184.159

The results of this study are consistent with the theories on the PSC particle growth continuum. Both type Ia and Ib PSCs are believed to start from stratospheric background aerosols, i.e., aqueous sulfuric acid solutions. Tabazadeh et al. (1994) suggested that liquid sulfate aerosols resulted in type Ib particle formation while frozen sulfate aerosols resulted in type Ia particles. Type Ia-enh clouds are characterized by nitric acid hydrate particles close to thermodynamic equilibrium (Tsias et al., 1999) and if the temperature is low enough, the type Ib clouds freeze to form type II clouds (Federico et al., 2001). Analysis of the SOLVE data supports these claims. The small amount of type II PSCs and many of the type Ib PSCs were observed on days with extremely cold temperatures (T < 188 K). In the image plots for January 23 in figure 8, a region of type II particles (purple) is found embedded inside a type Ib cloud (yellow). Figure 13 shows temperature versus δ_a for January 23 and illustrates the occurrence of type Ib and II PSC particles at the coldest temperatures.
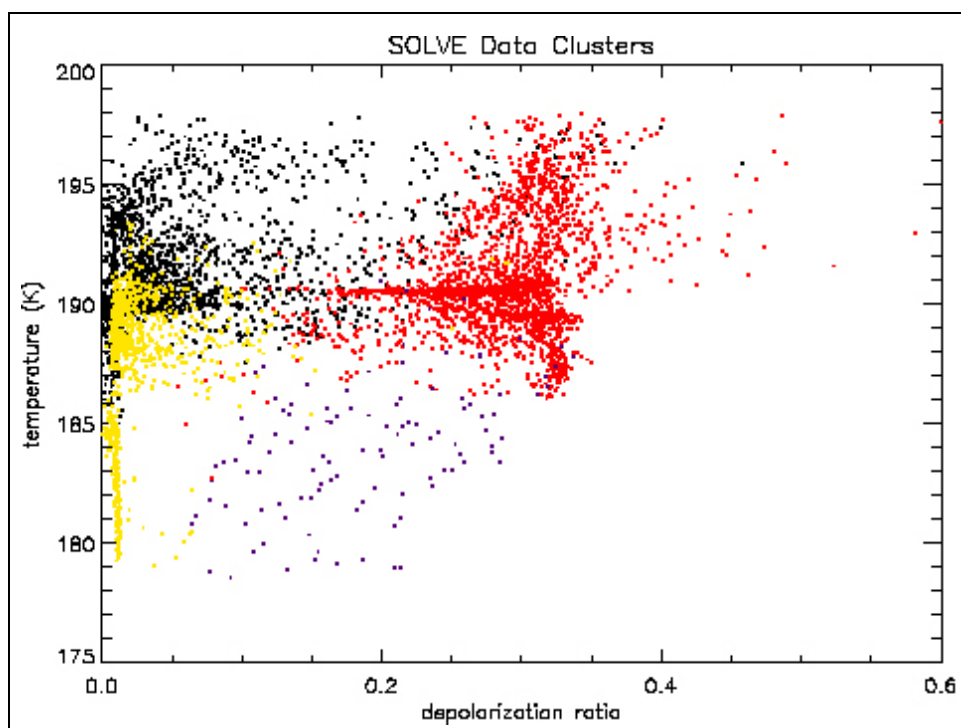
Figure 13. Temperatures at which the four clusters found on January 23 were observed.

Clouds were present in 7.27% of all nighttime data. The percentage of each PSC type found in the analyses is shown in table 10, and the frequency distributions for depolarization ratio and color ratio with the 532 nm backscatter coefficient are shown in figures 14 and 15. The concentrations of the aerosol liquid-solid mixture and the type II.

Table 10.   Percentage of the total number of nighttime data points for each PSC type found in the analysis.

PSC Type	Total Amount of Observations	Percentage
Aerosol liquid/solid mixture	8886	48.62
Ia	4714	25.80
Ib	2609	14.28
Ia-enhanced	1946	10.65
II	120	0.66

21
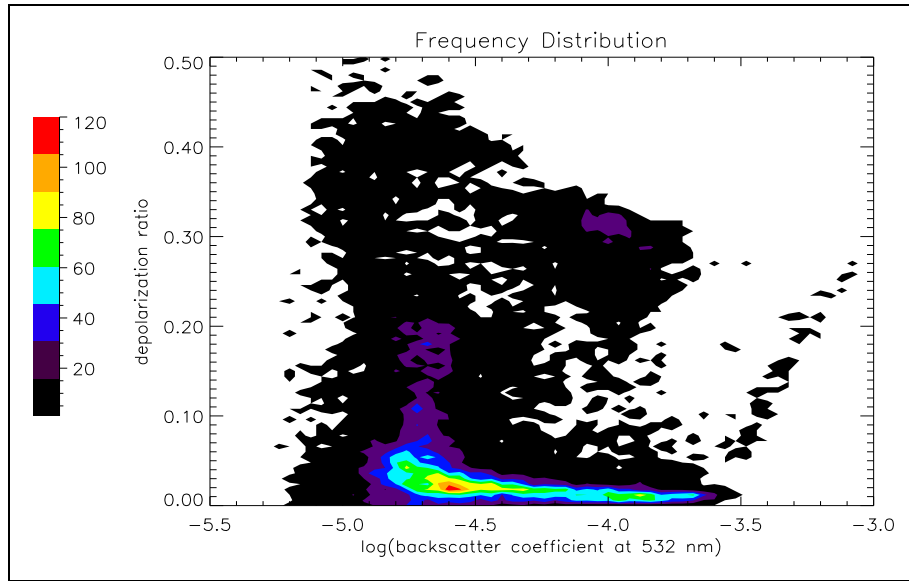
Figure 14. Frequency distribution for δ_a versus $\log(\beta_{532})$. (Colors correspond to total amount of observations in each bin.)
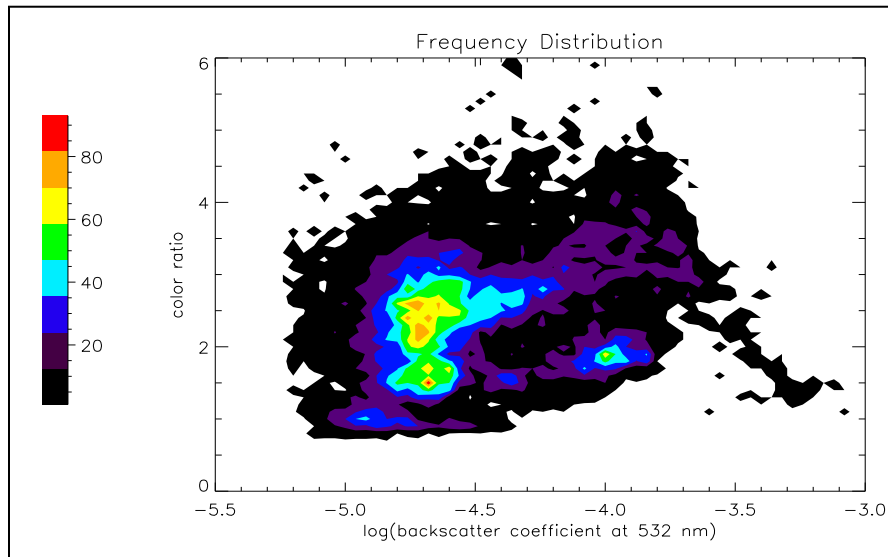
Figure 15. Frequency distribution for β_{532}/β_{1064} versus $\log(\beta_{532})$. (Colors correspond to total amount of observations in each bin.)

PSCs (48.62% and 0.66%, respectively, of the PSC data) are fairly typical of data taken in the northern hemisphere because the temperatures usually are not extremely cold long enough to allow for the formation of many Type II PSC clouds. Consistent with many other studies, the type Ia PSCs occurred most frequently.

5. Summary

A classical statistical technique was used to objectively identify PSC types and their corresponding characteristics with data acquired by the LaRC aerosol LIDAR during the SOLVE mission. Combinations of backscatter LIDAR measurements from which particle abundance, phase, and size can be inferred are used to classify the data and deduce PSC types. The variables used in the analyses are R_{532}, R_{1064}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}, and T. PCA showed that the bulk of the variance in the data set is attributed to scattering. Particle phase and size account for the second most weighted component of the analysis followed by temperature.

Cluster analysis was used to classify the clouds into mutually unknown groups, based on combinations of the derived variables from the PCA.

It was found that using R_{532}, δ_a, β_{532}, β_{1064}, β_{532}/β_{1064}, and T as input to the analysis resulted in the best clusters. When all 18,275 nighttime data points were used as input to the analysis, a liquid-solid aerosol mixture and types Ia and Ib were accurately identified.

Another cluster consisting of both type Ia-enhanced and type II particles was also found.

The small relative number of type II PSC events impeded the algorithm's ability to identify type II PSCs as a separate cluster.

Cluster analysis was also performed on two individual days of the SOLVE mission. January 23, 2000 contained 85.8% of the type II particles. When cluster analysis was performed on this day, the type II particles were recognized as a distinct cluster along with an aerosol cluster and types Ia-enhanced and Ib. The CA of March 5, 2000, which did not exhibit temperatures cold enough for the formation of type II (T ≤ 188 K), resulted in an aerosol cluster and clusters for types Ia and Ib.

The clustering algorithm also provided a measure of how well the clusters are defined, which can be used to determine the most natural number of clusters for the data set. This measure, referred to as the silhouette coefficient, proved to be insufficient for objectively identifying the number of PSC types present. This inability may be attributed to the PSC particle growth continuum discussed by Tabazadeh et al. (1994), Tsias et al. (1999), and Federico et al. (2001), which precludes a clear distinction between the optical properties of many of the PSC types. An additional factor that may contribute to the low silhouette coefficients may be the uncertainty in the LIDAR and temperature measurements.

The concentration of PSC particles in the data may affect the algorithm's ability to identify separate clusters. This may have been the case with the type II particles in this analysis. When the entire data set was clustered, there were only 176 data points with $R_{532} \geq 5$ of a total of 18,275. For all variable combinations, the algorithm failed to separate type II particles until at

least k = 6. For instance, for k = 5, the type II particles were grouped with the type Ia-enhanced because they have similar δ_a and β_{532}/β_{1064}. When the data from January 23, which contained 85.8% of the type II particles, were clustered separately, type II particles were successfully identified with the k = 4 partition. When we used only data with $R_{1064} > 2$ the number of data points for all the days decreased to 15,179 and the type II PSCs were identified with the k = 5 partition.

It is likely that the classification of type II PSCs can be improved with Antarctic data sets since they occur much more frequently because of the colder temperatures. This may reduce the effects of type II PSCs occurring in such limited numbers with respect to the other types of PSCs. This research may also benefit from the incorporation of additional variables that may help the algorithm distinguish between the PSC types. Using aerosol depolarization at both the visible and infrared wavelengths may enable the algorithm to identify the cloud types presented by Toon et al. (2000). It is also worth noting that this method can be used for analyzing large stratospheric LIDAR data sets such as the imminent satellite-based LIDAR measurements made by CALIPSO space-borne LIDAR.

6. References

Biele, J.; Tsias, A.; Luo, B. P.; Carsslaw, K. S.; Neuber, R.; Beyerle, G.; Peter, T. Nonequilibrium coexistence of solid and liquid particles in Arctic stratospheric clouds. *J. Geophys. Res.* **2001**, *106* (D19), 22,991–23,007.

Browell, E. V.; Butler, C. F.; Ismail, S.; Robinette, P. A.; Carter, A. F.; Higdon, N. S.; Toon, O. B.; Schoeberl, M. R.; Tuck, A. F. Airborne LIDAR observation in the wintertime arctic stratosphere: polar stratospheric clouds. *Geophys. Res. Lett*. **1990**, *17* (4), 385–388.

Carslaw, K. S.; Peter, T.; Clegg, S. L. Modeling the composition of liquid stratospheric aerosols. *Rev. Geophys.* **1997**, *35*, 125–154.

Federico, F.; Hauchecorne, A.; Knudson, B. Analysis of polar stratospheric clouds using temperature and aerosols measured by the Alomar R/M/R LIDAR. *J.Geophys. Res.* **2001**, *106* (D20), 24, 127–24, 141.

Felton, M. Classification of Polar Stratospheric Clouds using LIDAR Measurements from the SAGE III Ozone Loss and Validation Experiment, Master's thesis, 2003.

Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons, Inc., 1990.

Kent, G. S.; Poole, L. R.; McCormick, M. P.; Schaffner, S. K.; Hunt, W. H.; Osborn, M. T. Optical Backscatter Characteristics of Arctic Polar Stratospheric Clouds. *Geophys. Res. Lett*. **1990**, *17* (4), 377–380.

McCormick, M. P.; Chu, W. P.; Grams, G. W.; Hamill, P.; Herman, B. M.; McMaster, L. R.; Pepin, T. J.; Russell, P. B.; Steele, H. M.; Swissler, T. J. High-Latitude stratospheric aerosols measured by the SAM II Satellite system in 1978 and 1979. *Science* **1981**, *214*, 328–331.

National Research Council, The Atmospheric Sciences Entering the Twenty-First Century, National Academy Press, Washington, D.C., 364 pp., 1998.

Preisendorfer, R. W. Principal Component Analysis in Meteorology and Oceanography, Elsevier, 1988.

Poole, L. R.; McCormick, M. P. Polar Stratospheric Clouds and the Antarctic Ozone Hole. *J. Geophys. Res.* **1988**, *93*, 8423–8430.

Reichardt, J.; Tsias, A.; Behrendt, A. Optical properties of PSC Ia-enhanced at UV and visible wavelengths: Model and observations. *Geophys. Res. Lett.* **2000**, *27* (2), 201–204.

Struyf, A.; Hubert, M.; Rousseeuw, P. Clustering in an Object-Oriented Environment. *Journal of Statistical Software* **1996**, *1* (issue 4), 1–30.

Tabazadeh, A.; Turco, R. P.; Drdla, K.; Jacobson, M. Z.; Toon, O. B. A study of Type I polar stratospheric cloud formation. *Geophys. Res. Lett*. **1994**, *21* (15), 1619–1622.

Tabazadeh, A.; Toon, O. B. The presence of metastable HNO_3/H_2O solid phases in the Stratosphere inferred from ER 2 data. *J. Geophys. Res.* **1996**, *10 1*(D4), 9071–9078.

Toon, O. B.; Browell, E. V.; Kinne, S.; Jordan, J. An Analysis of LIDAR Observations of Polar Stratospheric Clouds. *Geophys. Res. Lett*. **1990**, *17* (4), 393–396.

Toon, O. B.; Tabazadeh, A.; Browell, E. V.; Jordan, J. Analysis of LIDAR observations of Arctic polar stratospheric clouds during January 1989. *J. Geophys. Res.* **2000**, *105* (D16), 20589–20615.

Tsias, A.; Wirth, M.; Carslaw, K. S.; Biele, J.; Mehrtens, H.; Reichardt, J.; Wedekind, C.; Weib, V.; Renger, W.; Neuber, R.; von Zahn, U.; Stein, B.; Santacesaria, V.; Stefanutti, L.; Fierli, F.; Bacmeister, J.; Peter, T. Aircraft LIDAR observations of an enhanced type Ia polar stratospheric clouds during APE-POLECAT. *J. Geophys. Res.* **1999**, *104* (D19), 23961–23969.

Distribution List

ADMNSTR
DEFNS TECHL INFO CTR
ATTN DTIC-OCP (ELECTRONIC COPY)
8725 JOHN J KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218

DARPA
ATTN IXO S WELBY
3701 N FAIRFAX DR
ARLINGTON VA 22203-1714

OFC OF THE SECY OF DEFNS
ATTN ODDRE (R&AT)
THE PENTAGON
WASHINGTON DC 20301-3080

US ARMY TRADOC
BATTLE LAB INTEGRATION & TECHL
DIRCTRT
ATTN ATCD-B
10 WHISTLER LANE
FT MONROE VA 23651-5850

SMC/GPA
2420 VELA WAY STE 1866
EL SEGUNDO CA 90245-4659

US ARMY INFO SYS ENGRG CMND
ATTN AMSEL-IE-TD F JENIA
FT HUACHUCA AZ 85613-5300

COMMANDER
US ARMY RDECOM
ATTN AMSRD-AMR W C MCCORKLE
5400 FOWLER RD
REDSTONE ARSENAL AL 35898-5000

US ARMY RSRCH LAB
ATTN AMSRD-ARL-CI-OK-TP TECHL
LIB T LANDFRIED (2 COPIES)
BLDG 4600
ABERDEEN PROVING GROUND MD
21005-5066

NASA LANGLEY RESEARCH CENTER
ATTN A H OMAR
MAIL STOP 401 A
HAMPTON VA 23681-2199

NASA LANGLEY RESEARCH CENTER
ATTN C A HOSTETLER
MAIL STOPE 420
HAMPTON VA 23681-2199

US GOVERNMENT PRINT OFF
DEPOSITORY RECEIVING SECTION
ATTN MAIL STOP IDAD J TATE
732 NORTH CAPITOL ST., NW
WASHINGTON DC 20402

EASTERN MICHIGAN UNIV
DEPT OF GEOGRAPHY AND GEOLOGY
ATTN T KOVACS
205 STRONG HALL
YPSILANTI MI 48197

HAMPTON UNIV
CENTER FOR ATMOSPHERIC SCI
ATTN M P MCCORMICK
21 TYLER ST
HAMPTON VA 23668

DIRECTOR
US ARMY RSRCH LAB
ATTN AMSRD-ARL-RO-EV W D BACH
PO BOX 12211
RESEARCH TRIANGLE PARK NC 27709

US ARMY RSRCH LAB
ATTN AMSRD-ARL-CI-E P CLARK
ATTN AMSRD-ARL-CI-ES M FELTON
(5 COPIES)
ATTN AMSRD-ARL-CI-OK-T TECHL
PUB (2 COPIES)
ATTN AMSRD-ARL-CI-OK-TL TECHL
LIB (2 COPIES)
ATTN AMSRD-ARL-D J M MILLER
ATTN IMNE-ALC-IMS MAIL &
RECORDS MGMT
ADELPHI MD 20783-1197